

**RE: Fw: IH Data fitting - Exponential vs LOESS**

**Brattin, Bill** to: Bateson.Thomas, Benson.Bob

08/07/2012 10:16 AM

From: "Brattin, Bill" <brattin@srcinc.com>

To:

Cc: Berry.David@epamail.epa.gov, Kopylev.Leonid@epamail.epa.gov, Christensen.Krista@epamail.epa.gov, DeVoney.Danielle@epamail.epa.gov, "Hilbert, Timothy (hilbertj)" <HILBERTJ@UCMAIL.UC.EDU>

---

Tom

We can discuss all of this in greater detail tomorrow, but here are a few responses that I can offer now:

<![if !supportLists]>a) <![endif]>?The ?LOD? (this is a serious misnomer...much better is to say ?analytical sensitivity?) is not an inherent constant such as 0.01.? Rather, the values varies between samples, depending on the volume of air drawn through the filter and the area of filter examined (see discussion in ?white paper? sent previously).? It is very possible to have values of 0.001 or even lower.? At Libby, we sometimes analyze samples to an analytical sensitivity of 0.00001 cc-1.

<![if !supportLists]>b) <![endif]>When the IH data are used untransformed to estimate the arithmetic mean, ?non-detects? should be evaluated as zero, not assigned some surrogate value (see white paper Section 7.2).? I am not quite sure what the file called ?Fiber Samples LOD Unadjusted 914 06132012? was used for, but fitting of the IH data was performed using zeros for ?non-detects?

<![if !supportLists]>c) <![endif]>In my opinion, a sampling duration weighted mean is not the right way to go.? Sampling duration is a relatively arbitrary value that is not related to worker exposure duration, and is also not related to confidence in the measured value.? Because long duration samples are generally collected when dust levels are low and short duration samples are generally collected when dust levels are high, a sampling duration weighted average would have a high risk of underestimating the impact of high exposures on workers.? I hope UC staff will be available tomorrow to weigh in on this, but I have not heard if they are available or not.

<![if !supportLists]>d) <![endif]>UC will have to provide the details and answer your questions on how the LOESS fitting was done...I only have a general overview.

\*\*\*\*\*

Bill Brattin  
SRC, Inc.  
999 18th Street Suite 1150  
Denver CO 80202  
Phone: 303-357-3121  
Fax: 303-292-4755  
e-mail: [brattin@srcinc.com](mailto:brattin@srcinc.com)

---

**From:** Thomas Bateson [mailto:Bateson.Thomas@epamail.epa.gov]

**Sent:** Tuesday, August 07, 2012 9:37 AM

**To:** Bob Benson

**Cc:** Brattin, Bill; David Berry; Leonid Kopylev; Krista Christensen; Danielle DeVoney

**Subject:** Re: Fw: IH Data fitting - Exponential vs LOESS

Bob,

Thank you for sending the LOESS plots. Before getting into the question of using the exponential or something else, we'd like to understand a few other things.

1) We have a file that you sent to Danielle called "Fiber Samples LOD Unadjusted 914 06132012". I have attached a copy. What do the "sample\_CONC\_F\_CC" = 0.001 represent? Danielle had said that 0.001 f/cc is too low to be measured. Is she right or wrong?

*(See attached file: Fiber Samples LOD Unadjusted 914 06132012 (from UC).xlsx)*

2) How were the concentration values for the non-detects generated (we think they were LOD/sqrt(2) )? Why they are so variable?

3) I saw that all the sampling duration data are here. I'm sorry that I had not appreciated that we had this data in hand. I compared the sampling-duration weighted arithmetic mean to the simple arithmetic average. Here are some of the results for track\_unload:

Year	Avg	Wt-Avg
1976	107.3	100.8
1977	4.52	0.485
1978	0.75	0.143
1979	7.27	6.45
1980	52.9	52.9
1981	0.26	0.014
1982	0.06	0.06

So, some years it makes no difference but for others like 1977, the difference can be large.

4) I think that we should be using sampling-duration weighted arithmetic mean (and geo mean) and NOT the simple averages. Do you concur? Do you want to check with UC folks?

5) Until we settle this, it makes no sense to look at the exponential plots versus the LOESS. However, when we do come back to this issue, we'd like to see both the exponential and LOESS plots on the same graph.

We would also like a description of the methods used to fit the LOESS. What kind of smoother was used? Maybe a normalized kernel? I'm assuming that the "smooth" was that which optimized the fit. The curve for "background" looks very bumpy.

Something about the statistics you sent struck me as odd. The LOESS for trionizing looks pretty smooth and if this is the optimal fitting "smooth" parameter, then the LOESS shouldn't have used many DF to fit. The stats show this to be true (d.f. 538 vs 540). So how can the MSE be more than 10-fold different? Seems odd. Maybe check all those numbers again. Perhaps the reasons will be more evident when both plots (exp and LOESS) are on the same graph and we understand the methods used).

Thanks,

Tom

Thomas F. Bateson, ScD MPH  
Epidemiologist  
Effects Identification & Characterization Group  
EPA/ORD/NCEA  
1200 Pennsylvania Ave. NW (Mail Code 8623P)  
Washington, DC 20460

Phone: 703-347-8570

▼ Bob Benson---08/02/2012 02:00:15 PM---Tom, Here is another file that compares fitting LOESS v Exponential. Distribute as need to the rest

From: Bob Benson/R8/USEPA/US  
To: Thomas Bateson/DC/USEPA/US@EPA  
Cc: [brattin@srcinc.com](mailto:brattin@srcinc.com), David Berry/R8/USEPA/US@EPA  
Date: 08/02/2012 02:00 PM  
Subject: Fw: IH Data fitting - Exponential vs LOESS

---

Tom,  
Here is another file that compares fitting LOESS v Exponential. Distribute as need to the rest of the NCEA Team.

Region 8 agrees with the recommendation to use the exponential fit to establish the JEM. Do you concur?

----- Forwarded by Bob Benson/R8/USEPA/US on 08/02/2012 11:52 AM -----

From: "Brattin, Bill" <[brattin@srcinc.com](mailto:brattin@srcinc.com)>  
To: Bob Benson/R8/USEPA/US@EPA,  
Cc: Tim Hilbert <[HILBERTJ@UCMAIL.UC.EDU](mailto:HILBERTJ@UCMAIL.UC.EDU)>, David Berry/R8/USEPA/US@EPA  
Date: 08/02/2012 07:43 AM  
Subject: IH Data fitting - Exponential vs LOESS

---

Bob

UC has completed fitting the IH data using LOESS.  
The attached files present graphs and give goodness of fit summary statistics.  
In brief, based on statistics alone  
LOESS is best for track

Exponential is best for trionizing and unload  
There is little difference for background

UC has indicated they think it is probably best to stick with the exponential  
fit for all, and I am inclined to agree.

Let's sent this to Tom and see what he thinks.

\*\*\*\*\*

Bill Brattin

SRC, Inc.

999 18th Street Suite 1150

Denver CO 80202

Phone: 303-357-3121

Fax: 303-292-4755

e-mail: brattin@srcinc.com<<mailto:brattin@syrres.com>>

*(See attached file: winmail.dat)(See attached file: LOESS v exponential.xlsx)(See attached file:  
scan.pdf)*